DOCUMENT RESUME

ED 237 880 CG 017 135

AUTHOR Feldman, Jack M.
TITLE The Practice, Practicality, and Prospects of Training for Performance Appraisal.
PUB DATE Aug 83
NOTE 23p.; Paper presented at the Annual Convention of the American Psychological Association (91st, Anaheim, CA, August 26-30, 1983).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Methods; *Job Performance; Outcomes of Education; *Personnel Evaluation; Psychometrics; State of the Art Reviews; *Training Methods
IDENTIFIERS *Performance Appraisal

ABSTRACT

Historically, training for performance appraisal has focused on the same issues as instrument development--the reduction of psychometric errors in ratings. Efforts were centered around teaching people to use rating scales properly. A review of the literature shows these programs met with mixed success. While a meta-analysis of these data are premature, several hypotheses may be drawn: (1) knowledge of the job in question is more important than rating skills; (2) observational skills are important in real-world ratings; (3) the purpose and context of ratings are as or more important to accuracy than the training itself; and (4) accuracy should be the primary goal of training. Training for performance appraisal is far from universal. Most training efforts in actual use involve learning how to use a particular form or system. One possible training method to improve accurate evaluations involves the use of multiple performance examples, such as videotape, to represent multiple levels of accomplishment. Little systematic knowledge exists about the mechanics of implementing a theoretically-based appraisal system. It is necessary to understand how the appraisal system functions in the operation of the organization. Considerations of equity, of the multidimensionality of job performance, or the cost of more refined observations may make more sophisticated measurement impossible to achieve. Reliable, valid measures that provide accurate determination of two or three levels of performance are an advance over biased assessment of five, or six or more. (JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237880

The Practice, Practicality, and Prospects of Training for Performance Appraisal

Jack M. Feldman

University of Florida

Presented at the American Psychological Asscociation
Anaheim, California

August, 1983

CG 017135

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jack M. Feldman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Practice, Practicality, and Prospects of Training for Performance Appraisal

Jack M. Feldman

Department of Management and Administrative Sciences

University of Florida

Gainesville, FL

As some of you may know, much of my work over the last few years has been concerned with theoretical analyses of the performance appraisal process, in terms of some fairly abstract principles based on theories of cognition and social perception. More recently, (Feldman, 1981) I have tried to turn these ideas into guidelines for the development of appraisal instruments and training procedures for appraisers (who, after all, are really the measuring instruments).

Today I'm going to try to deal with the practicalities of training for appraisal; I'd like to consider some factors relevant to the development of theoretically based instruments and the use of appropriate training. I will not consider employee feedback per se; since the success of feedback depends first on accurate assessment, the latter is my concern here. I'd also like to suggest some areas of common concern for the laboratory oriented, theoretically minded researcher and the observation-oriented, organizationally minded researcher/practitioner, consideration of which might benefit both.

Historically, training for performance appraisal has focussed on the same issues as instrument development - the reduction of "psychometic errors" in ratings. The frequency and size of halo errors, leniency/-stringency biases, contrast effects, and so forth were the dependent variables of interest. Efforts were centered around teaching people to use rating scales "properly", i.e. to avoid halo, contrast, etc. These programs

met with mixed success. Brown (1968) used a training program emphasizing practice with the rating scale, a discussion of rating errors, and an emphasis on "trait differentiation", finding reduced halo (increased inter-scale variance) for trained peer raters on a set of six trait scales. Borman (1975) found that brief training in the recognition and avoidance of halo error reduced its magnitude in the rating of specially constructed videotaped performance vignettes, though reliability decreased as well. In agreement with Brown (1968), Bernardin & Walter (1977) found that training and familiarization with behavioral expectation scales reduced halo and leniency error in instructor ratings by students. This training apparently improved the observational skills of students, as well as focusing their attention on the performance dimensions covered by their Behavioral Expectation Scale.

Bernardin (1978) found that a one-hour training session was more effective than a five-minute session in reducing leniency and halo in student ratings; however, the training effect disappeared after a few months. Support for the proposition that longer, more detailed training sessions (especially when used in conjunction with behaviorally-anchored rating scales) are more effective also comes from Wexley, Sanders, and Yukl (1973). Their effective interviewer training session involved discussion of a job's requirements and applicant qualifications, a detailed evaluation guide, videotaped examples of good, bad, and average performers together with rating feedback and discussion of psychometic error. They were successful in eliminating contrast errors in the rating of videotaped stimuli. A similar six to eight hour program designed by Latham, Wexley & Purcell (1975) likewise featured instruction in observational skills, discussion of errors, feedback and active participation in learning to eliminate errors.

They found that, six months later, the workshop training eliminated contrast, halo, and similarity, errors in comparison to a control group, and recency effects compared to a "discussion training" condition. Videotaped interviewees were the stimuli. Borman (1979) likewise found that an intensive workshop reduced halo in the ratings of videotaped stimuli, but did not improve accuracy.

Ivancevich (1979) used an even more intensive (three-day) training procedure, compared to a three-day discussion group and a no-treatment control. The intensive training involved videotaped performance examples and feedback to managers. Intensive training was superior to the discussion and control conditions in reducing halo and leniency in actual administrative ratings six months after training, but the effect on halo was reduced after 12 months. Warmke & Billings (1979), likewise conducting research in an organizational setting, compared the effectiveness of shortened discussion training patterned after Latham, et al. (1975) with lecture, participation in graphic rating scale construction, and a control group. On experimental ratings, participation in scale construction and lecture were most effective in reducing psychometric errors, and the lecture group produced marginally greater interater reliability. Interestingly, on halo effect measures, there was a significantly greater degree of error when ratings were made for administrative rather than research purposes.

To further confuse things, Bernardin & Pence (1980) found that training to recognize and avoid rating errors, which included examples, discussion, and feedback, was effective in reducing halo and leniency compared to a control group and a second group trained in the dimensions of the job in question. However, the trained group was less accurate than either of the other groups in rating the hypothetical stimulus vignettes.

These results correspond broadly to those obtained by Borman (1975, 1979), who found no effects of either a short (1975) or long (1975) training program on the accuracy with which subjects rated written vignettes or videotaped stimuli. Apparently, training in the avoidance of psychometic error changes rating behavior (e.g. leads to greater between-dimension rating variance); but, as Bernardin & Pence (1980) concluded, the new response sets may distart the representation of performance by incorrectly lowering scores and removing "true" halo (Cooper, 1981).

Recognizing that a meta-analysis on these data is premature, and realizing the dangers of drawing conclusions from narrative (and brief) reviews, may any useful hypotheses be drawn from these studies? I think yes.

They are:

1. Knowledge of the job in question is as or more important than rating skills. Participation in scale development probably teaches one about the job. A useful program should teach appraisers what behaviors to look for as well as how to translate observations into numbers on paper.
2. Observational skills are important in real-world ratings, which must be made on the basis of events occurring over long periods of time.
3. The purpose and context of ratings are as or more important to accuracy than the training itself.
4. As stated by Latham & Wexley (1981) and Cooper (1981), "psychometic errors" in actual administrative ratings may not be errors at all, but may reflect

> real intercorrelation among job dimensions (halo), or high or low levels of work-group performance (leniency-stringency). It follows that accuracy should be the primary goal of training.

That observational accuracy can be taught is supported by Thornton & Zorich (1980) as well as Bernardin & Walter (1977). The importance of rating purpose is underscored by Zedeck & Cascio (1982), who found that training had no effect on the evaluation of hypothetical supermarket employees, but purpose of the evaluation (merit raise vs. employee development or retention) did. In particular, the "raise" decision resulted in less differentiation among the hypothetical employees.

Finally, the idea that observational accuracy is related to accuracy in appraisal is supported by Murphy, Garcia, Kerkar, Martin, and Balzer (1982), in their study of the evaluation of videotaped lecture performance. While their study did not focus on training effects, it seems reasonable that, if individual differences in observational accuracy are related to differences in appraisal accuracy, then training to improve the former should also improve the latter.

The program most clearly related to the hypotheses above is that of Latham & Wexley (1981). This involves an intensive workshop focussing on a succession of psychometric errors (e.g., similarity, halo) with videotaped stimuli and behavioral feedback. The program also trains observational skills. They report that this type of program is successful in actual practice, in that the training applied to a group of supervisors improved criterion reliability and validity sufficiently so that a previously "invalid" selection battery successfully predicted the "new" criterion scores. Unfortunately, it is not known just what aspects of the training

contributed to the improvement. Given the data presented above, though, I would hypothesize that the observational skills component was largely responsible for the increase in validity. A study comparing training in the several components of the program with a group taking the complete training course (and including no-treatment and placebo groups) would properly test this hypothesis.

Their program does not deal with training in the job itself. Apparently, they assume that supervisors are sufficiently knowledgeable and competent to render this step superfluous. They seem to rely on knowledge of the principles of observation, judgement, and rating to transfer to the job in question.

In conclusion, I think the evidence is consistent with the contention that training in the avoidance of psychometric error, in and of itself, is not helpful. The people who become more accurate appraisers most probably learned what kinds of behaviors to observe, and how to observe and recall them in the context of a valid conception, or schema, of the job. This conclusion agrees with Borman's (1979) regarding the goals of training.

Training in Practice

Training for performance appraisal is far from universal, whether in the private or the public sector. Estimates of its frequency range from 75% to less than 25% (De Vries, Morrison, Shullman & Gerlach, 1981). In an attempt to discover the nature and content of current training efforts, I searched several practitioner's journals from the most recent issue through 1978. The results were disappointing, to say the least.

While articles about performance appraisal are not infrequent, very few mention training at all. In one recent survey of appraisal practices (Teel, 1980), no mention was made of training for appraisers. Some do mention the

need for training (e.g., Wells, 1982) to assure consistent application of the appraisal system. One author (Beaulieu, 1980) recommends at least 40 hours of training for appraisal, in addition to appraisal monitoring and followup systems. But neither specifies exactly what the training is supposed to include, nor how it is to be accomplished.

Only one paper (Robinson & Robinson, 1978) dealt with the nature of training. The authors discussed Performax^{R}, a modelling-based program designed to teach managers how to conduct goal-setting, performance feedback, and appraisal. Managers are taught to establish specific goals and standards and provide daily feedback. They are apparently not taught how to establish standards or assess one's level of success in meeting them. These vital skills go largely unmentioned in the practicioner-oriented literature. They may be learned via modelling of one's own supervisor, implying a large chance component in skill development. Skill training may also be part of some consultants' programs; if so, the practice seems far from universal.

One large-scale performance appraisal program (Gomez-Mejia, Page, & Tornow, Notes 1,2) includes a more ambitious training program. The training their system provides includes computer-based instruction in the requisite company policies and the use of the appraisal system, but also includes a 12-hour workshop on the appraisal process itself. This workshop seems very similar to the training described by Latham, et al. (1975) though the actual training content is not discussed. Their extensive workshop and computer-based training do allow the possibility of training and feedback in observation, encoding, and judgment skills.

The most defensible conclusion can draw from this effort is that most training efforts in actual use involve learning how to use a particular form or system (e.g., Haynes, 1978). In contrast, the training programs

recommended in the applied academic literature focus on the elimination of psychometric error in ratings and the development of observational skills. The fact that most appraisal systems in use involve some form of training indicates that a recognized need for training exists. It is the responsibility of the academic researcher to develop more useful forms of training, and to demonstrate that usefulness in ways that lead to the adoption of our best programs. What follows is an outline of the form I believe training should take.

Training Based on Theory

So far, we have seen two forms of training; the first derived from empirical work with minimal theoretical background, the second from the popularization of that work as well as the earlier "form-centered" research on appraisal. What improvements can the newer theory-centered approaches promise the practitioner?

One thing that should be remembered is that any new approach is going to contain elements of previous practice. Just as a 14th century archer did not need Newtonian mechanics to hit the target, so good empirical research and practice may be valid without extensive theoretical underpinnings. The role of theory here is to explain what is observed and to improve practice by pointing out relationships not previously considered. But there are a lot of steps between theory and technology.

First of all, we must differentiate between observation, encoding storage, recall, and evaluation or rating. Accurate evaluations depend first of all on the observer attending to important and relevant behaviors, then encoding or categorizing these appropriately, and recalling them when needed. In my system, this depends very much on the appraiser's cognitive structure or category system as well as transient factors influ-

encing his or her available categories. It follows from this that part of the success of previous training programs is due to this focus on the behavioral definition of explicit job dimensions, either through scale development, lecture, or discussion. These become part of the "job schema" or category/prototype system used in appraisal. To the extent trainees learn to recognize relevant behaviors automatically, an important component of accuracy is added to the appraisal process.

As one might expect, those who are more experienced and better at the job are more valid raters (Landy & Farr, 1980). An interesting study by Levy (1960; reported in Campbell, Dunnette, Lawler, & Weick, 1970) showed that high-performing accounting supervisor's evaluations of subordinates correlated with subordinate intelligence, while poorer-performing supervisor's ratings correlated with clerical aptitude. This at least makes plausible the notion that one's "implicit theory" (schema) of the job influences one's ratings, and that training should cover important aspects of the subordinate's job itself.

Second, we must teach the translation of events into judgments. What actions are regarded as good or poor, and <u>how</u> good or poor? We have seen that contrast effects may bias such judgments, and that training in scale use may alleviate them. This is the point at which instrument-centered training and feedback and anchoring stimuli (as in behavioral expectation scales) are most useful. At this point, raters should be taught to avoid bias caused by job-irrelevant categorization (e.g., race, sex) or overall evaluative impressions, so as to reduce <u>illusory</u> halo (e.g., Nathan & Lord, 1983).

One possible training method would involve the use of multiple performance examples (e.g. videotaped performances, or products of

performance) sampled so as to represent multiple levels of accomplishment on each of the schema-given dimensions of job behavior. The trainee could evaluate relevant performance dimensions at a computer console and get immediate feedback as to the fit of his or her judgements to an "ideal" evaluation model. After initial training, interpolated task activity could be introduced between observation and rating, so that both short and long-term memory for (and encoding of) behaviors could be assessed. Such training could include examples of job performance by people differing on job irrelevant dimensions (e.g. age, race, sex) so that potential biases could be "trained out" of the rating response. The method is similar to the procedures used in concept attainment studies. It may be modified as appropriate for different types of tasks, as discussed below.

A third point, taken from the cognitive perspective, is that different evaluation instruments and different types of training are appropriate for evaluating the performance of different types of jobs. This echos De Vries, et al. (1981), though it was developed independently (Feldman, 1981). My thesis in that earlier paper is taken from Hammond (1981): there is a continuum of cognitive tasks, anchored at one extreme by the "analytic" and at the other by the "intuitive". The midpoint is represented by the "quasi-rational" task. An analytic task is represented by a mathematics problem, or mechanical assembly; there is an unequivocal standard for judging its performance, and the process of its performance is accessible to consciousness. The intuitive task is exemplified by the building of scientific theory, or creativity in the arts. The process of solution is not entirely accessible to consciousness, and there are multiple standards of evaluation that can be applied. The quasi-rational tasks contains elements of both -

the job of architect, for example, has analytic elements (structural specifications) and intuitive elements (artistic merit).

Lower level jobs are often analytic, and may be defined by behaviorally-anchored scales and objective criteria. The evaluator must be trained to be an observer and recorder; presumably the value of each performance dimension may be discovered by cost accounting or some validated estimation procedure (e.g., Bobko, Karren, & Parkington, 1983), so that the appraiser does not have to scale the behaviors.

Upper-level jobs are often quasi-rational or intuitive in nature. The intuitive component requires the appraiser to either choose or develop an appropriate task schema and then use it to evaluate performance. This, in my opinion, is the theoretical basis for any usefulness in MBO and similar procedures, as recommended by De Vries, et al. (1981). The evaluator must be trained not only in observation and evaluation, but in the multiplicity of possible approaches to the job in question. In this case, job experience or reputation does not guarantee adequate evaluation; "scholarship" is needed. The critic may not be able to act or paint, but must know a lot about acting or painting.

Finally, for the quasi-rational task, both kinds of skills are needed, as appropriate to the task dimension. That is, analytic task dimensions must be evaluated using one kind of form, with a particular type of training. Intuitive task dimensions require different forms, and different training. Finally, the two types of evaluations must be weighted and combined to produce an overall judgement appropriate to the decision in question.

It is appropriate to note here that cognitive/developmental psychologists have recently begun to stress the importance of previous knowledge to

new learning (Siegler, 1983). Children (and, I am willing to bet, adults as well) are said to learn by experiencing exceptions to previously held rules for encoding and inference, whereupon new rules are adopted and tested.

It is well-known that experts use different categories and rules than novices for encoding and inference. It follows that one important function of applied research must be the discovery of the categories and inference rules that are presently used, in order to aid training in new category systems and inference rules where necessary. The training method discussed earlier can accomplish this.

What evidence is there for the usefulness of this approach? Frankly, none. Borman's (1979) finding that different rating formats were more accurate for different jobs, though limited,is at least consistent with these ideas. It is also encouraging that others have come up with somewhat similar ideas. I do however, have some ideas about how to test these notions.

The basic strategy is one of construct validation. Job dimensions should be analyseable in terms of well-validated ability and/or personality constructs. If an appraisal procedure is in fact more accurate or less biased, performance as measured by that procedure ought to correlate with measures of the relevant abilities or dispositions and not with others. These correlations should be higher than those obtained with other, equally reliable, appraisal procedures. Furthermore, to the extent that performance as measured in Job 1 depends on dimensions also common to Job 2, appraisals on Job 1 should be valid predictors of Job 2 performance - more valid than other predictors. Appropriate evaluator training and appropriate evaluation instruments ought to improve real-world predictability; inappropriate training should reduce the obtained correlations. We should also find that

experienced, competent incumbents rate subordinates and peers as expected on the basis of independently-derived job schemata, and longitudinal studies should show the development of these schemata over time. For intuitive tasks, the ability to generate and use multiple schemata should exist in experts, regardless of their preference for one particular schema or another.

Contextual Moderators of Training Effectiveness

So far, I have been dealing with training in a vacuum, as if the evaluator was free to give any rating he or she desired, and as if accurate appraisal was the only goal of the appraisal system. Neither assumption is generally true.

As Ilgen and I discussed in our 1983 paper, performance appraisal is an integral part of organizational functioning. Training people to use a system that does not fit the realities of their organization is at best a waste of time for all concerned. Consider the military and civil-service performance appraisal systems. In the military, an elaborate set of forms and procedures are used for "form's sake", but are essentially meaningless. "Real" evaluations are communicated by a series of key words, known through experience and word-of-mouth. In the civil service, the system is so formal and legalistic as to prevent meaningful personnel actions, and any attempt to change the appraisal system requires revamping the entire structure.

The private sector would seem to offer more flexibility, but even here the requirements of accurate appraisal are often subordinate to individual and group agendas. Organizational politics may require promotions for certain subordinates, regardless of their relative qualifications. "Keeping the peace" in a work group may require equal raises for all, again regardless of performance differences. "Merit-based" pay may require high

evaluations for all when salary budgets are bountiful, and low evaluations when they are lean. Such factors are far more powerful than the ideal of appraisal accuracy, and will exist regardless of training. It is therefore necessary to influence the entire structure and reward system of the organization if we expect even the best possible appraisal system to function properly.

Benefits and Costs of Appraisal Systems and Training

It should be possible to justify the necessary organizational changes on a purely economic basis. Data on the increased predictability of job performance resulting from improved appraisals can be used to project economic benefit to the firm, as done by Schmidt, Hunter, McKenzie & Muldrow (1979) for selection devices. Likewise, improvements in job satisfaction have consequences for turnover and other costly behaviors, and a better appraisal system may contribute importantly to satisfaction. Accurate appraisals also allow pay to be used in a maximally motivating manner (e.g., Lawler, 1981) improving both morale and productivity. We do not have accurate estimates of the financial outcomes of such interventions, but their estimation is certainly feasible.

Other potential benefits derive from the current legal environment. Recent court decisions have established that performance criteria must be standardized, objective, and job-related, and based on a formal job analysis. Appraisers should be trained in the system, which itself should pertain to well-defined standards of behavior or performance. In addition, performance criteria used for promotion decisions must meet the same standards as other selection devices. A program of instrument development

and training based on cognitive theory, if supported by both laboratory and field-study results, meets these standards.

Finally, and perhaps most importantly, employees can benefit - from reductions in role ambiguity, from clear standards for reward and advancement, from recognition of the truly outstanding performers, and from a system which admits less personalistic and group-centered bias.

Costs of such a system are perhaps more difficult to estimate. Estimating direct costs of development, of course, is not a great problem-consulting fees, man-hours, computer time, training time, and so forth can be handled easily. Other costs-time lost due to the change of established power relationships, anxiety, initial dissatisfaction, etc. will be extraordinarily difficult to quantify.

Perhaps we may estimate these by looking at similar large-scale organizational changes-job enrichment, for example - as a way of setting upper and lower bounds on costs. Early, small-scale implementation of such programs (e.g., in a few plants of a large corporation) using quasi-experimental techniques may also help cost/benefit estimates. If, as De Vries, et al. (1981) and Teel (1980) state, appraisal systems undergo frequent revisions, the incremental costs of an innovative system can be more easily justified.

Problems of Implementation

In general, little systematic knowledge exists about the mechanics of implementing a theoretically-based appraisal system. We know that acceptability of any new system is important in practice, and that a system that presupposes a fundamental change in organizational relationships (at least sometimes) is likely to be unacceptable to some. It is also likely to be resisted and sabotaged, regardless of high level support.

In order to understand how to introduce a new appraisal system and make it effective, we most first understand how the appraisal system functions in the operation of the organization. This is a task for the observation-oriented researcher. Systematic, quantitative observational data are needed, not anecdotes or case studies; these data should focus on the characteristics of formal and informal appraisal systems across organizations of different types, of different degrees of success, in different cultures, under different economic constraints. I take as a fundamental assumption that systems, both formal and informal, evolve to serve some purpose. We need to discover the systems that exist, and their purposes. How, for example, is employee performance represented in Japan, in both traditional organizations and more Western ones? How does this system differ from that in other Oriental locales (e.g., Taiwan, Hong Kong)? How does the system differ by industry? By the degree of "industrial democracy" as found in many European nations? The more we know about the kinds of systems that exist, their precursors and their ramifications, the better we can plan for changes in our own system.

At the individual level, we should investigate the nature of category systems and inference rules that actually exist. We may, for example, discover that similar kinds of schemata and rules are commonly used in organizations with more valid appraisal systems, or that expert appraisers use similar rules regardless of organization. This may be one more ramification of the generality of cognitive skill.

Finally, we should face the possibility that we may not be able to refine our measures of job performance past the point of identifying two or three levels of contribution. Considerations of equity, of the multidimensionality of job performance, of the cost of more refined observations and

so forth may make more sophisticated measurement impossible to achieve. If so, we can still make sure that the measures we use are reliable and valid as the state of the art will allow, and take comfort in the fact that accurate determination of two or three levels of performance is an advance over the unreliable or biased assessment of five, six, or more.

References

1. Gomez-Mejia, L.R., Page, R.C., & Tornow, W.W. Key steps in a successful performance appraisal program and its relationship to other personnel functions. Manuscript submitted for publication. Department of Management and Administrative Sciences, University of Florida, Gainesville, Florida, 32611.

2. Gomez-Mejia, L.R., Page, R.C., & Tornow, W.W. Key steps in developing, implementing, and monitoring an effective performance appraisal system. Manuscript submitted for publication, Department of Management and Administrative Sciences, University of Florida, Gainesville, Florida, 32611.

References

Beaulieu, R. An easier look at performance appraisal. Training & Development Journal, 1980, 34 (10), 56-59.

Bernardin, H.J. Effects of rater training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 1978, 63, 301-308.

Bernardin, H.J., & Pence, E.G. The effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 1980, 65, 60-66.

Bernardin, H.J., & Walter, C.S. Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 1977, 62, 64-69.

Bobko, P., Karren, R., & Parkington, J.J. Estimation of standard deviations in utility analyses: An empirical test. Journal of Applied Psychology, 1983, 68, 170-176.

Borman, W.C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 1975, 60, 556-560.

Borman, W.C. Format and training effects on rater accuracy and rater errors. Journal of Applied Psychology, 1979, 64, 410-421.

Brown, E.M. Influence of training, method, and relationship on the halo effect. Journal of Applied Psychology, 1968, 52, 195-199.

Cooper, W.H. Ubiquitous halo. Psychological Bulletin, 1981, 90, 218-244.

De Vries, D.L., Morrison, A.M., Shullman, S.L., & Gerlach, M.L. Performance appraisal on the line. New York: Wiley-Interscience, 1981.

Feldman, J.M. Training and instrumentation for performance appraisal: A perceptual-cognitive viewpoint. Proceedings of the annual meeting, American Institute for Decision Sciences, Nov. 1981, Vol. 1, 339-341.

Hammond, K.R. Principles of organization in intuitive and analytical cognition. (Report #231, Center for Research on Judgement and Policy). Boulder, Colorado: University of Colorado, 1981.

Haynes, M.E. Developing an appraisal program. Personnel Journal, 1978, 57, 66-67, 104, 107.

Ilgen, D.R., & Feldman, J.M. Performance appraisal: A process focus. In Cummings, L.L., & Staw, B.M. (Eds.) Research in Organizational Behavior (Vol. 5). Greenwich, Conn. JAI Press, 1983.

Ivanceivich, J.M. Longitudinal study of the effects of rater training on psychometic error in ratings. Journal of Applied Psychology, 1979, 64, 124-131.

Landy, F.J., & Farr, J.L. Performance rating. Psychological Bulletin, 1980, 87, 72-107.

Latham, G.P., & Wexley, K.N. Increasing productivity through performance appraisal. Reading, Mass.: Addison-Wesley, 1981.

Latham, G.P., Wexley, K.N., & Pursell, E.D. Training raters to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 60, 550-555.

Lawler, E.E., III. Pay and organization development. Reading, Mass.: Addison-Wesley, 1981.

Levy, S. Supervisory effectiveness and criterion relationships. Paper presented at the annual convention of the American Psychological Association, Chicago, Ill., May 1960. Reported in Campbell, J.P., Dunnette, M.D., Lawler, E.F. III, & Weick, K.E. Jr. Managerial behavior, performance and effectiveness. New York, McGraw-Hill, 1970.

Murphy, K.R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W.K. Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 1982, 67, 320-325.

Nathan, B.R., & Lord, R., G. Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 1983, 68, 102-114.

Robinson, J.C., & Robinson, L.C. Modelling techniques applied to performance feedback and appraisal. Training and Development Journal, 1978, 32 (1), 48-53.

Siegler, R.S. Five generalizations about cognitive development. American Psychologist, 1983, 38, 263-277.

Teel, K.S. Performance appraisal: Current trends, persistent progress. Personnel Journal, 1980, 59, 296-301,316.

Thornton, G.C. III, and Zorich, S. Training to improve rater accuracy. Journal of Applied Psychology, 1980, 65, 351-354.

Warmke, D.L., & Billings, R.S. Comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. Journal of Applied Psychology, 1979, 64, 124-131.

Wells, R.G. Guidelines for effective and defensible performance appraisal systems. Personnel Journal, 1982, 61, 776-782.

Wexley, K.N., Sanders, R.E., & Yukl, G.A. Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 1973, 57, 233-236.

Zedeck, S., & Cascio, W.F. Performance appraisal decision as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 1982, 67, 752-758.